

VIP 2024-VAA

Group Project Update 9/3/2024



Meeting Agenda

- Last week, we reviewed research paper summaries, discussed data labeling, and potential research questions
- This week, we looked for interesting research papers and possible literature reviews, got comfortable with using OpenMMPose, and labeled antelope data

Semi-Supervised Keypoint Detection Paper

Title: Semi-Supervised Keypoint Localization

Authors: Olga Moskvayak, Frederic Maire, Feras Dayoub, Mahsa Baktashmotlagh

Publication year: 2021

Publication venue: ICLR

What's new in this paper:

The research paper focuses on improving performance of semi-supervised keypoint detection with limited labeled data, by adding semantic keypoint representations to be used by a keypoint classification network (KCN) onto existing keypoint localization networks (KLN).

Model architecture: HRNet-32

Data: MPII, LSP, CUB-200-2011, ATRW

Method: The paper uses semi-supervised learning, with keypoint heatmaps and semantic keypoint representations.

Visual Transformer Pose Estimator

Title: ViTPose++: Visual Transformer for Generic Body Pose Estimation

Authors: Yufei Xu, Student Member, IEEE, Jing Zhang, Senior Member, IEEE, Qiming Zhang, Student Member, IEEE, and Dacheng Tao, Fellow, IEEE

Publication year: 2023

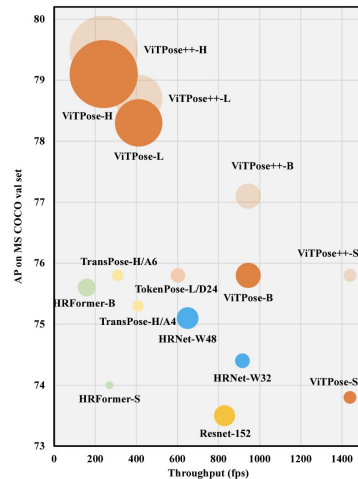
Publication venue: IEEE Transactions on Pattern Analysis and Machine Intelligence

What's new in this paper:

The research paper focuses on developing the basis for a foundational pose estimation model for all body types(human and animal), using a modular and simple visual transformer to make the model more generalizable.

Data: MS COCO, AI Challenger, OCHuman, MPII, ImageNet, COCO-Wholebody, AP-10K and APT-36K

Method: The ViTPose++ model implements a simple encoder, by splitting the image into varying numbers of patches of RGB streams and applying attention and feed forward layers to them. Then it has multiple different decoder models from a simple decoder, to a task specific decoder. The different decoders can be utilized for top-down and bottom-up pose estimation. Each decoder performs a form of heatmap regression. Masking encodings in the pretraining resulted in better performance on more datasets(more generalizability between human and animal pose estimation).



Next Steps:

- Continue coming up with research questions
- Continue reading more papers on pose and keypoint detections
- Continue to experiment with the different models and adjust hyperparameters to check differences in outputs

Personal Progress

Josh Mansky

- Annotated 100 images with labels (set 1)
- Re-read through the OpenPose paper to clear up confusion on how it performs Bipartite matching for grouping key points together, and its results and points of failure
- Searched for animal pose estimation literature reviews
- Read and analyzed ViTPose++ paper
 - Learned how the transformer works
 - Learned how simple visual transformer works
- Possible Research Questions:
 - How do we determine the behavior of animals(antelopes) from poses extracted from short camera trap videos of them?

Claire Kim

- Annotated 100 images with labels (set 2)
- Read the papers, *Semi-Supervised Keypoint Localization* and *ViTPose++: Visual Transformer for Generic Body Pose Estimation*
- Need help with trying to open mmpose
 - When running command from “Course Guidelines” pptx, get this error:

```
⊙ (base) vip24_shared@deadcat:~/mmpose$ python demo/image_demo.py sample_images/01150167_0360_crop00_detections.jpg configs/animal_2d_keypoint/topdown_heatmap/animalpose/td-hm_hrnet-w32_8xb64-210e_animalpose-256x256.py hrnet_w32_animalpose_256x256-1aa7f075_20210426.pth --out-file ./sample_images/output1_animalpose.jpg --draw-heatmap
Traceback (most recent call last):
  File "/home/vip24_shared/mmpose/demo/image_demo.py", line 5, in <module>
    from mmcv.image import imread
ModuleNotFoundError: No module named 'mmcv'
```


Medha

- Annotated 100 images with labels (Set 4)
- Read Open Pose and Animal Pose Research Papers
- Found and analyzed Semi-Supervised Keypoint Localization Paper
- Familiarized with running OpenMMPose
- Brainstormed some potential research questions
 - What is the best ratio of labeled to unlabeled data for semi-supervised keypoint detection with limited data?
 - Does semantic keypoint consistency work to improve the accuracy of HRNet on antelope data?
 - Does semantic keypoint consistency work to improve the accuracy of RTMPose on antelope data?
 - By how much do transformation consistency constraints improve performance and does it increase generalization?

Armaan

- Annotated 100 images with labels (set 5)
- Read AP-10k paper
- Analyzed paper on existing animal pose specific kit: DeepPoseKit
 - <https://docs.google.com/document/d/1tPwH0hlaGe7ZSG-j-bxuX7b0frpeocz4F2JcdE9bV9U/edit?usp=sharing>
- Possible Research Questions:
 - What is the optimal balance between synthetic and real data for enhancing keypoint detection accuracy in underrepresented animal species?
 - How does leveraging deep learning-enhanced bounding box frameworks affect the robustness of keypoint detection in dynamically changing environments?
 - What effect does combining semantic masks with bounding boxes have on keypoint detection accuracy in cluttered scenes?

Parth Thakre

- Annotated 100 images with labels (set 6)
- Read Animal Pose and Open Pose papers and glanced over Semi Supervised Keypoint paper.
- Possible Research Questions:
 - Can using animal identification help create more precise bounding boxes to improve keypoint detection accuracy?
 - How can we adjust keypoint detection techniques (e.g., by modifying different labeling techniques for different keypoints) to enhance accuracy?

Shaan

- Labeled 100 images
- Read the paper that introduced HRNet, and learned about multi-scale fusion
- Possible Research Questions:
 - To what extent is including labeled keypoint data on one species beneficial for extracting key points on another species?